

Pradeep Pujari

Home: 510-894-1468

<https://www.kaggle.com/ppujari>

Email: ppujari@hotmail.com

<https://linkedin.com/in/ppujari>

<https://twitter.com/ppujari>

<https://github.com/ppujari>

SUMMARY:

I am an innovative and result-driven deep learning engineer proficient in Machine Learning, Computer Vision, NLP, and scalable web architecture. I am actively seeking a challenging role where I can leverage technical expertise and problem-solving skills to develop cutting-edge software solutions for customers.

CORE COMPETENCIES:

- Extensively worked on Machine Learning, **Deep Neural Networks**, **Computer Vision**, Large Language Models (LLM), Reinforcement Learning, Natural Language Understanding (NLU), Graph Neural Networks, Transformer Model and its variants Knowledge Graphs, and Information Retrieval.
- Hands-on experience with Big Data technologies including Hadoop, Spark and Cloud Computing.
- Experience in architecting, designing, and coding scalable e-commerce platforms.
- Proficient in **data structures, algorithm design and complexity analysis**
- Hands-on experience in RDBMS, NO-SQL database Cassandra, Mongo, Vector database FAISS
- Good at Object Oriented Design principles, Design Patterns, middleware components.
- Hands-on experience in RAG ML pipeline, prompt engineering, CoT, LangChain, DSPy
- Hands-on experience Data quality control, Versioning, Standardization, Data set generation
- Enjoy Mentoring and fostering team cohesion, work effectively in cross-functional teams!

TECHNICAL SKILLS:

Programming Languages

Java, Python, C++, Pandas, NumPy

ML/Deep Learning Framework

Scikit-Learn, TensorFlow, Keras, **PyTorch**, OpenVINO, Gym, GAN

Machine Vision Tool Kit

OpenCV, CNN, Faster R-CNN, YOLO, **CLIP, BLIP, GLIDE**

Graph Neural Network

NetworkX, PyTorch Geometric

MLOps

Apache Ray, SageMaker, Kubernetes, FeatureStore - FEAST

GenAI/NLP Tool Kit

spaCy, Core NLP, **BERT, T5, GPT, BART, LangChain, DSPy**

Search Science Technology

Lucene-Solr, Elastic Search, Nutch, **Neural IR**, WordNet, LLMs

Distributed Computing

Redis, Hadoop, PySpark, Hive, ZooKeeper, **Amazon Bedrock**

Networking & Operating System

Edge Caching, UNIX internals, Tomcat, Flask

SQL/NoSQL/Vector Database

Cassandra, Mongo, Hbase, Oracle, PostgreSQL, Vector DB FAISS

Distributed Streaming

Kafka, KSQL, Faust, Apache Storm, REST API

OPEN-SOURCE CONTRIBUTION:

Lucene-Solr Project, OpenAI

- Created multi-agent collaboration projects such as training agent pairs for tennis @Udacity.
- Implemented DDPG paper in PyTorch, evaluated deep RL models.
- Designed and coded Neural Machine Translation @Udacity
- Road lane detection, path planning @Udacity Self driving course.
- **LLM - Detect AI Generated Text @Kaggle My model accuracy was 0.851**
- Worked on Few-shot Question Answer LLM Model and Identifying Age-Related Conditions
- **Vision Language Models** – Stable Diffusion, BLIP, Interrogating CLIP@Huggingface

Technology: **DDPG-Actor-Critic**, OpenAI Gym, Unity ML-agents, **PyTorch, BERT, LLaMA**

WORK EXPERIENCE:

EleutherAI – AI Research Scientist (Part Time)

May2024-Present

- Collaborated with researchers to develop evaluation metrics, run performance evaluation pipelines, and debug ML models.
- Applied **DPO** to medical vision-language models to align model behavior with human preferences.

GenAI Engineer – KP Digital:**Feb 2022-Present**

- Designed and Developed Medical Image to Prompts - Stable Diffusion Model Using Open Source LLM with Prompt and instruction tuned with RLHF.

Natural Language Processing for Literature Mining project:

- Applied NLP techniques for knowledge extraction from scientific literature and databases. Developed text mining tools for literature curation, knowledge discovery, and hypothesis generation. Identified and anonymized PII data.
- Designed and built scalable evaluation systems that automated model assessments, integrating seamlessly with CI/CD pipelines.

Technology: PySpark, Azure, Docker, Kubernetes, PyTorch, T5, BART, MedAlpaca, LLaMA2**Software Engineer-Meta:****Aug 2021 - Feb2022**

- Benchmarked and optimized **data and model parallelism** for a large-scale **SparseNN Ad Ranking ML model** on cutting-edge AI accelerators and chips, improving performance and efficiency for Meta's high-traffic advertising infrastructure.
- Developed and enhanced a **personalization model** that tailored ad content to individual user preferences, leveraging real-time data and optimizing for both relevance and engagement, which led to improved user experience and ad effectiveness.
- Applied hyperparameter tuning, quantization, and QLoRA to optimize memory usage and accelerate inference while maintaining model accuracy.
- Collaborated with cross-functional teams to integrate new hardware, utilizing **Python** and **CUDA** for AI model acceleration and GPU optimization.

Technologies: Hyperparameter Tuning, Quantization Techniques, Personalization Models, QLoRA, Python, CUDA Programming, PyTorch**Principal NLP Engineer – CVSHealth (contract)****Apr 2021 - Aug2021**

- Medical Imaging for Tumor Detection- involves several imaging modalities, advanced image processing techniques, and machine learning algorithms to accurately identify and diagnose tumors.
- Collected Dataset: ChestX-ray14, ISIC, The Cancer Imaging Archive (TCIA), **BRATS**: Brain Tumor Segmentation Challenge dataset. Deployed the model in a clinical setting.

Technology: Python, PyTorch, **OpenCV**, **scikit-image**, **CNN**, AWS, MLflow, Amazon Bedrock**Principal Machine Learning Engineer – Oto Analytics****Feb 2021 - Jul2021**

- Designed, developed, and implemented **Feature Store -Feast**, ML pipeline.

Technology: Python, Redis, Apache Airflow, Apache Kafka, AWS, MLflow, **Amazon Bedrock****Machine Learning Architect - ServiceNow (contract)****Aug 2020 - Jan2021**

- Worked on Intent extraction (identify contradiction) using **BERT**, Natural Language Inference
- Architected and Built Chatbot conversational AI system for Help Desk with **RASA NLU Lib**

Technology: Python, Transformers, AWS, Mongo, Nutch, RASA – NLU, SquAD, BART, T5**Principal Machine Learning Engineer - Kohl's Innovation Lab****Jan 2018 - Jul2020**

- Implemented a real-time object detection system non-rigid tracking using FPGA-based hardware acceleration with pre-trained deep learning models YOLO, SSD
- This project involves FPGA programming, interfacing with camera modules, and optimizing the inference pipeline for real-time performance.
- Store Shopping Intent Analysis models for **people counter**, **Age and Gender identification**.
- **Porting model in mobile and Camera edge devices**

Technology: Python, pyTorch, OpenCV, spaCy, YOLO, 3D CNN, Fast R-CNN, OpenVINO

Staff Machine Learning Engineer - Walmart Labs**May 2015 - Mar 2017**

- Built ML models for attribute extraction such as detecting brand, color, size etc. with
 - A. Supervised Learning B. Sequence Labelling and CRF C. CNN approach.
 - Relation extraction, Semantic Parsing, and knowledge graph
- Implemented Neural Network model for Item Matching System based on title and description.
 - Used pre trained Word embedding – **Glove**-contextual word representation.
- Worked in distributed data pipeline to orchestrate raw item JSON through a series of **micro services** producing a sellable item.
- High-Performance Computing (HPC) workloads: Large-Scale Machine Learning Training: Built distributed machine learning frameworks to train deep neural networks on massive datasets.

Technology: Python, Scikit Learn, TensorFlow, Keras, ZooKeeper, Kafka, Storm, Cassandra, Mongo**Senior Architect - Angie's List****Mar 2012 - Mar 2014**

- Designed and built **Geospatial Search** that increased revenue and customer retention by 25%.
- Designed and programmed Query understanding and rewriting with a heuristic Ranking module.
- Integrated **NLP** annotator tools into Solr using Apache UIMA for **Phrase extraction**.
- Worked on JVM performance tuning and set up search infrastructure cloud cluster.
- Implemented **Multi Objective Session based Recommender System** for Deals, Coupons

Technology: Java/J2EE, Lucene-Solr, Python, Multitask Ranking, Kea, Ling Pipe, NER, Elastic Search**Staff Software Engineer - Walmart Labs Search and Platform Team****Jun 2010 - Mar 2012**

- Designed Sentiment Analysis Network: Product Reviews data set – sentiment aware tokenizer.
- Built Meta-Search Engine with Apache Carrot, Enhanced content with annotation engine-UIMA.
- Object Tracking in video, Handwritten digit recognition, Hand gesture recognition, pose detection

Technology: Scikit Learn, SentiWordNet, SGD, Sentiment Treebank, Word2Vec, OpenCV, Flask, REST**Tech Leader – Search Science - Macys.com****May 2001 - May 2010**

- Implemented automatic IR Evaluation System. **MS-MARCO** document ranking dataset, Real-Time Indexing, Search Ranking Algorithms, Auto Suggest component, Sponsored Search
- Implemented web crawler-NUTCH secure and static web pages, Faceted Search and Browse
- New Ranking algorithm for Search Relevancy Tuning, Query Understanding and rewrite,
- Successfully completed customer segmentation with k-means clustering
- Implemented Deep Personalization ranking mechanism based on a user's search and click history.

Technology: Java J2EE, Tomcat, Lucene-Solr, Vector Space Model, BM25, Learn to Rank, Elastic Search**EDUCATION****MS in Computer Science and Applications** - National Institute of Technology, India

Jun 1988

B. Sc (Physics Hons) - Berhampur University, India

Jun 1984

Data Science Specialization – Johns Hopkins University

May 2015

AI Specialization 3 Semester course certification– Stanford University

July 2009

PROFESSIONAL DEVELOPMENT**Udacity Project: Home Service Robot****May–Dec2020**

Programmed a Home service Robot that maps environment and navigate to pickup and deliver objects.

Technology: C++, ROS, Monte Carlo Localization, Path Planning**Udacity Project: Self Driving Car Nano degree****Oct – Dec2017**Completed: a. Advanced Lane Lines detection b. Vehicle Detection c. Extended Kalman Filter
d. Traffic Sign Classifier e. behavioural cloning f. controls MPC e. PID control, Simulation and Testing**Technology: C++, PyTorch, OpenCV, PCL (Point Cloud Library), Mask R-CNN****PUBLICATIONS AND PRESENTATIONS:**

Coauthor of Book “Practical Convolutional Neural Network” – Packt Publisher

Feb 2018

Presented “Sentiment Analysis using Solr” at Sentiment Symposium, San Francisco

May 2013

Presented Paper “Detecting Cyber bullying instances” NLU course

Apr 2022

Pradeep Pujari

Cell: 510-396-9057

3 /3